

2016-02

The logic-bias effect: The role of effortful processing in the resolution of belieflogic conflict

Howarth, S

<http://hdl.handle.net/10026.1/12528>

10.3758/s13421-015-0555-x

Memory & Cognition

Springer Science and Business Media LLC

All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.

The logic-bias effect: The role of effortful processing in the resolution of belief–logic conflict

Stephanie Howarth¹ · Simon J. Handley¹ · Clare Walsh¹

Published online: 21 September 2015
© Psychonomic Society, Inc. 2015

Abstract According to the default interventionist dual-process account of reasoning, belief-based responses to reasoning tasks are based on Type 1 processes generated by default, which must be inhibited in order to produce an effortful, Type 2 output based on the validity of an argument. However, recent research has indicated that reasoning on the basis of beliefs may not be as fast and automatic as this account claims. In three experiments, we presented participants with a reasoning task that was to be completed while they were generating random numbers (RNG). We used the novel methodology introduced by Handley, Newstead & Trippas (*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 28–43, 2011), which required participants to make judgments based upon either the validity of a conditional argument *or* the believability of its conclusion. The results showed that belief-based judgments produced lower rates of accuracy overall and were influenced to a greater extent than validity judgments by the presence of a conflict between belief and logic for both simple and complex arguments. These findings were replicated in Experiment 3, in which we controlled for switching demands in a blocked design. Across all three experiments, we found a main effect of RNG, implying that both instructional sets require some effortful processing. However, in the blocked design RNG had its greatest impact on logic judgments, suggesting that distinct executive

resources may be required for each type of judgment. We discuss the implications of our findings for the default interventionist account and offer a parallel competitive model as an alternative interpretation for our findings.

Keywords Reasoning · Working memory · Deductive reasoning · Individual differences · Problem solving · Decision making

Many theorists in the fields of reasoning and decision making have argued that thinking is characterized by the action of two distinctive cognitive systems. Such accounts are referred to as *dual-process theories* and have been developed in a number of areas, including learning (e.g., Reber, 1996), attention (Shiffrin & Schneider, 1977), reasoning (Evans, 2003), decision making (Kahneman & Frederick, 2002), and social cognition (Chaiken & Trope, 1999). Associated with these distinct cognitive systems are different processing characteristics. Evans, (2009) suggested that these processes can be classified as Type 1—fast, automatic, unconscious, and effortless by nature—and Type 2—slow, controlled, conscious, effortful, and demanding on the working memory. In other areas of research, these processes have been variously referred to as *intuitive* or *rational*, *automatic* or *deliberative*, or *heuristic* or *analytic* (Evans, 2009; Sloman, 1996; Stanovich, 1999). For the purposes of this article, we employ Evans's (2009) classification and terminology of *Type 1* and *Type 2* processes throughout.

According to the default interventionist dual-process account of reasoning (detailed in Evans & Stanovich 2013), Type 1 processes are assumed to yield default, automatically cued responses that may be based upon beliefs, learned associations, heuristics, or stereotypes. These responses can be resisted or intervened on by conscious and deliberative higher-order reasoning processes (Type 2); however,

✉ Stephanie Howarth
stephanie.howarth@plymouth.ac.uk
Simon J. Handley
S.Handley@plymouth.ac.uk
Clare Walsh
clare.walsh@plymouth.ac.uk

¹ Department of Psychology, Plymouth University, Drake Circus, Plymouth, Devon PL4 8AA, UK

producing an alternative response often requires cognitive effort and the inhibition of Type 1 outputs (Handley, Capon, Beveridge, Dennis, & Evans, 2004). As a result, our initial intuitive response to a problem is often accepted (Stanovich, 2009). It is worth noting that biased responding can also arise through Type 2 processes when there is limited engagement of reflective processing; furthermore, Evans & Stanovich (2013) have argued that there is always some degree of Type 2 processing in response generation, even if it is minimal. However, under all default interventionist accounts of reasoning, a prototypical example of frugal reasoning that has established itself as a significant phenomenon and is often cited as supportive of dual-process accounts is the “belief-bias” effect.

Belief bias reflects the tendency for individuals to generate and authenticate a conclusion on the basis of their beliefs rather than of the logical structure of an argument. Take, for example, the following problem as an illustration of the conflict between belief- and logic-based reasoning:

*All plants need water
Roses need water
Does it follow that roses are plants?*

The instructions given in reasoning studies generally require the participants to determine whether the conclusion follows logically from the premises. In this example, the conclusion does not logically follow, but it is consistent with our beliefs. According to default interventionist accounts, our beliefs are cued by default, often leading to high endorsement rates for these types of conclusions.

It is claimed that individuals often experience a cognitive struggle between what they believe is true about the world and what is logically entailed. This is often reflected in reduced reasoning accuracy and increased latencies on problems in which beliefs and logic conflict, even when participants are given explicit instructions to reason on the basis of the logical structure of the argument (Evans, Barston, & Pollard, 1983). It has been claimed that prior beliefs are extremely difficult to override and compete for control over the responses made (Evans, 2003). Consequently, the conflict between whether or not individuals believe the conclusion or agree with its logical validity is often resolved in favor of what they find believable, which has been considered a reliable support for a *default*-based system in reasoning (Evans, 2007).

A number of findings appear to support the notion that the believability of a conclusion is automatically available, and thus preempts logical analysis (Evans et al., 1983; Stanovich, 1999). Syllogistic-reasoning studies carried out under limited time appear to increase the occurrence of belief bias (Evans & Curtis-Holmes, 2005). Increasing cognitive load by adding a demanding secondary task to be completed alongside the reasoning task also increases the belief-bias effect (De Neys,

2006). Furthermore, those with higher cognitive ability are able to resolve conflict more readily and to resist a belief-based response (Stanovich & West, 1998). This suggests that informants who lack the resources to review and uphold an alternate argument in their working memory, or an argument that is decontextualized in nature, are more likely to lead to an automatic, Type 1 output.

Default-processing accounts tend to emphasize the preemptive impact that prior beliefs have on our decision-making processes, often leading them astray (Evans, 2006). In contrast, recent research has indicated that reasoning on the basis of beliefs may not be as fast and automatic as these accounts claim. For example, belief bias on simple reasoning tasks is unaffected by a speeded problem presentation (Evans, Handley, & Bacon, 2009), and measures of belief bias in reasoning do not consistently correlate with cognitive capacity measures (Newstead, Handley, Harley, Wright, & Farrelly, 2004). Furthermore, developmental research has shown that among preadolescent children, biased responding on reasoning tasks is more, rather than less, common in participants who score higher on measures of fluid intelligence and working memory (Morsanyi & Handley, 2008). One explanation for these findings is that reasoning about beliefs may often be an effortful process, relying on the activation of relevant knowledge and integration with a novel problem structure in order to arrive at a judgment informed by relevant beliefs. In addition to this, there is evidence that outcomes normally attributed to Type 2 processes, such as reasoning on the basis of logical structure, can be achieved by automatic, fast (i.e., Type 1) processes (Rader & Sloutsky, 2002) and can be drawn without awareness (Leo & Greene, 2008). The claim that logical reasoning can be accomplished automatically is also supported by recent research that has suggested that participants automatically detect logic–belief conflict, even when this is resolved in favor of belief-based responding (De Neys, 2012).

A common feature of the studies that have examined the impact of beliefs on reasoning is that the instructions typically require participants to assume that the premises are true and to draw conclusions on the basis of the premises’ logical validity. One possibility is that belief bias arises, not because of the default generation of knowledge-consistent responses, but because of a failure to appropriately apply the instructional set. In a series of experiments, Handley, Newstead, and Trippas (2011) examined the accuracy and time course of processing when participants were instructed to respond on the basis of either logical validity *or* the believability of a given conclusion. The believability of the conclusion was evaluated with respect to empirical beliefs—for example,

*All zabs can walk.
Whales are zabs.
Therefore, whales can walk.*

This novel methodology allowed a direct test of the default interventionist account, which would predict that when participants were asked to make judgments of conclusion believability, as opposed to conclusion validity, responses should both be more accurate and take less time. Their findings consistently showed that belief-based judgments take significantly longer than those made under logical instructions and that judgments relating to strongly held beliefs could themselves be undermined if they were inconsistent with the logical structure of an argument (Handley et al., 2011). These findings were interpreted as showing that certain logical inferences are accomplished relatively automatically and are immune to the influence of beliefs. This pattern is *contrary* to what would be expected if belief-based judgments were based on fast, automatic Type 1 processes and logic-based judgments were based on slow, deliberate Type 2 processes, indicating that the distinction between Type 1 and Type 2 processes is far from clear.

An alternative dual-process theory, contrasting with the default interventionist framework, is the parallel-processing model. This model is common in social psychology (Chaiken & Trope, 1999) but has also been applied to the judgment and decision-making literature (Sloman, 1996). According to the parallel competitive account, Type 1 and Type 2 processes are initiated at the same time and operate in competition with each other. The process that completes first prompts a response that may need to be inhibited in favor of a more appropriate, yet less rapid, one. However, inhibition of the first response will depend on the instructions, the resources available, and the time it takes for each process to complete. Handley and Trippas (2015) have recently developed a new parallel-processing account that assumes that logical and belief-based judgments draw upon both Type 1 and Type 2 processes, but that the extent of competition depends on the complexity of each type of judgment, and hence the extent to which an alternative available response must be resisted in favor of additional processing. This account is a useful framework for understanding Handley et al.'s findings, if we assume that certain reasoning problems prompt a rapid logical response (as opposed to a default, belief-based response) that requires inhibition when one is presented with a belief-based instructional set (Handley et al., 2011).

These more recent findings (Handley et al. 2011) suggest that on certain tasks, belief-based judgments should be just as affected by speeded tasks and secondary loads as logical judgments are. This proposal is the foundation for the three experiments presented in this article. The present research implements the original methodology of Handley et al. (2011), which allowed us to evaluate the degree to which logic and beliefs interfere with one another while participants engage in a secondary task. The default interventionist model, which highlights the sequential nature of processing—in which

Type 1 precedes Type 2 processing (Evans, 2006, 2011)—predicts that performance on belief-based judgments (early responses) should be unaffected by the secondary task. The parallel competitive model, on the other hand (see also Handley & Trippas, 2015), predicts that belief judgments should be affected by the secondary task as much as logic judgments are. We examined this prediction across three experiments, which include both simple and more complex arguments, as well as individual difference measures associated with effortful processing.

Random number generation (RNG) has been one of the most commonly used methods to explore the role of the central executive component of working memory in cognitive tasks. RNG is thought to draw on a number of executive functions, including “inhibition” and memory “updating” (Miyake et al. 2000). It has been well established that the generation of random responses is disrupted by and impedes simultaneously performed, attention-demanding secondary tasks (Baddeley, 1986; Brown, Soliveri, & Jahanshahi, 1998), including syllogistic reasoning (Gilhooly, Logie, Wetherick, & Wynn, 1993). Our aim was to examine whether belief-based reasoning would be disrupted by a secondary load that interferes with Type 2 processes, such as the engagement in random number generation.

Experiment 1

In Experiment 1, we examined whether applying a secondary load would have a greater impact on belief or logic judgments on simple modus ponens (MP) arguments—in particular, when the believability and logical validity of a conclusion were in conflict. If logic judgments are more resource-demanding, in accordance with the default interventionist account, we would expect them to be influenced more than belief-based judgments would be; alternatively, if belief-based judgments also engage Type 2 processing, then they should also be affected by RNG.

Method

Participants

A total of 90 participants took part in Experiment 1: 56 were women and 36 were men (mean age = 31 years). Seventy-six of the participants were paid £4 each, and 14 were awarded course credits for their time.

Design and materials

A 2 (Instruction) × 2 (Problem Type) × 2 (Load) mixed design was used, in which participants were randomly assigned to the RNG (load) or control (no load) group. Each participant was

presented with the same 64 conditional reasoning problems in MP format. Of these problems, 32 were presented under belief instruction and 32 under logical instruction (Instruction Type factor). In each of these sets were 16 conflict problems (with conclusions that were valid and unbelievable or invalid and believable) and 16 no-conflict problems (with conclusions that were valid and believable or invalid and unbelievable). In each case, eight problems required a “yes” response or a press of the “s” key on the key board, and eight problems required a “no” response or a press of the “k” key.

Reasoning task The reasoning task was presented on a computer screen; Table 1 demonstrates the type of questions presented to each participant. Argument 1 shows that the conclusion of the argument logically follows from the premise but is in conflict with beliefs. In Argument 3, the conclusion is believable but the conclusion does not follow logically. For Arguments 2 and 4, the logical validity and believability of the conclusion coincide, and hence there is no conflict. The conclusion to each of the invalid arguments (Arguments 3 and 4 in Table 1) involves the denial of the consequent clause. In order to avoid an association between validity and explicit negation, antonyms were used in the conclusion clause.

We created 16 scenarios of the kind shown in Table 1, with four items within each scenario: two conflict and two no-conflict items. To counterbalance the problems, one group of participants were presented with a conflict and a no-conflict item from each scenario under belief instructions and the remaining two items under logic instructions, whereas the second group were presented with the same scenarios and items under different instructions (see Appendix A for the full set of problems used). Thus, participants received equal numbers of belief-based and logic-based questions, equal numbers of logically valid and invalid conclusions, and equivalent numbers of “yes” and “no” responses.

Table 1 Examples of the different types of modus ponens (MP) arguments used in Experiments 1–3

Conflict Arguments	No-Conflict Arguments
Argument 1 If it is raining then the street is dry. It is raining. (A) The street is dry. Logic: ✓ Belief: X	Argument 2 If it is raining then the street is wet. It is raining. (C) The street is wet. Logic: ✓ Belief: ✓
Argument 3 If it is raining then the street is dry. It is raining. (B) The street is wet. Logic: X Belief: ✓	Argument 4 If it is raining then the street is wet. It is raining. (D) The street is dry. Logic: X Belief: X

For each problem, the first premise was presented on the computer screen for a total of 3,000 ms and remained on the screen to ensure that participants had enough time to read and understand the first premise before the second premise, conclusion, and response options appeared together. Before the experiment started, four practice trials were presented, which covered all four argument types (see Table 1).

Random number generation The RNG task was based on the methodology used by Miyake et al. (2000). Participants were instructed to say aloud a number between 1 and 9 every second for the entire duration of the experiment. They were instructed to continue generating random numbers while solving the reasoning problems. To ensure that they understood the concept of “random,” they were given the following example:

Suppose you had written the numbers 1–9 on pieces of paper and put them in a hat. You take out one piece of paper, call out the number and return it to the hat. Then you reach for another piece of paper and do the same thing. The series of numbers you call out in that way should be random. (Horne, Evans, & Orne, 1982)

Before the start of the experiment, they were told to keep the generation of numbers continuous; although it was tempting to pause while reading the questions, they had to keep generating a number every second to the sound of the metronome beat. They were also advised that if they accidentally went beyond the number range 1–9, they should try *not* to pause, but should continue on.

Procedure

Participants were tested in maximum groups of four and were randomly assigned to one of two conditions: RNG (load) or the control group (no load).

They were tested in partitioned booths behind closed doors to keep vocalized distractions to a minimum. Participants in both the RNG and control groups wore closed-cup earphones to reduce background noise, but only the RNG group listened to a metronome beat that was set to click every second.

Participants were informed that they would have to complete 64 reasoning problems and were instructed to answer either according to their beliefs or according to logic. Belief instructions emphasized the requirement to answer in relation to their knowledge of what is true in the world and to indicate whether the conclusion was believable or unbelievable. They were then presented with following example:

If you finish your drink then the glass will be full.
Suppose you finish your drink.
Does it follow that the glass will be full?
s) BELIEVABLE k) UNBELIEVABLE

The correct answer according to beliefs is UNBELIEVABLE because, based upon your knowledge of the world, you know that if you finish your drink then the glass will be empty; therefore, the conclusion is unbelievable.

The logic instructions asked participants to assume each statement was true (even if in reality it was not true) and to indicate whether the conclusion followed validly from the preceding sentences. They were then presented with the same example under logic instructions:

If you finish your drink then the glass will be full.

Suppose you finish your drink.

Does it follow that the glass will be empty?

s) VALID k) INVALID

In this particular example, the correct answer according to logic is INVALID, because the first premise states that “if you finish your drink then the glass will be full,” and supposing that you “finish your drink,” you must logically conclude that your glass will be full. This is why the conclusion “the glass will be empty” is logically invalid.

In the RNG group, each participant had a dictaphone placed in front of the keyboard to capture the articulation of numbers. The control group only had the reasoning task instructions to adhere to, but the RNG group had separate instructions for the secondary task. Response latencies were logged from the presentation of the full problem until a response was given.

Analyzing randomness The RGCalc program (Towse & Neil, 1998) was used to analyze the randomness of the numbers generated by each participant in the high-load condition. The program produces many different indices needed to examine randomness, but the three main measures used here were taken from Towse and Valentine (1997) and are as follows:

Redundancy: A score of 0% means each response alternative is given with equal frequency, whereas 100% means the same response has been selected too often throughout—for example, 2,2,2,2.

RNG= random number generation: This measures how often a response alternative follows another. The closer that RNG is to 100%, the more predictable the pair sequence.

Adjacency: This measures a specific type of sequential pairing, in contrast to the RNG analysis—for example, the commonality of 1,2 or 3,2 (neighboring pairs on the number line). A score of 0% means that no neighboring pairs were presented, whereas 100% means all successive

responses were adjacent number values. In Experiment 1, the three measures were used as a check to ensure that participants engaged in the secondary task. A score of 50% or higher on two out of the three randomness indices was set as the criterion for eliminating participants not sufficiently engaging executive resources, although no participants were eliminated on these grounds. The average scores for the randomness indices were 3% for redundancy, 50% for RNG, and 40% for adjacency. Three participants were removed from the analysis because they took longer than 2 s per number generation.

Results

We analyzed the full data set before removing the data from 20 participants: three for taking longer than 2 s to generate numbers, and 17 for scoring below chance on the conflict problems (i.e., less than 50% accurate), suggesting that they were applying an inappropriate instructional set to the questions (Handley et al. 2011). This gave us totals of $N = 35$ in the RNG group and $N = 35$ in the control group. We focus on the reduced data throughout this article, although the pattern of findings from the full and reduced data sets did not differ meaningfully. All of the accuracy data were arcsine-transformed to improve the homogeneity of the variances and to control for the impact of ceiling effects. Analysis of the latency data was performed on correct responses only, and any missing data were replaced with the overall cell mean for that question type (i.e., belief conflict, belief no-conflict). The latencies were then stripped by two standard deviations to remove outliers and ensure a more even distribution of the means. The tables throughout this article present response latencies with outliers removed and percentage accuracy scores, prior to transformation. For this experiment, see Table 2. Missing and stripped data accounted for no more than 4% of the overall data.

A 2 (Instructions: belief/logic) \times 2 (Problem Type: conflict/no conflict) \times 2 (Load: RNG/control) mixed-design analysis of variance (ANOVA) on the accuracy scores revealed a main effect of problem type, $F(1, 68) = 73.274$, $p < .001$, $\eta_p^2 = .519$, showing poorer performance on conflict items (80% vs. 94%). We also observed a main effect of instructions, $F(1, 68) = 16.436$, $p < .001$, $\eta_p^2 = .195$, with logic judgments producing the most accurate scores (90% vs. 84%). There was also a main effect of load, $F(2, 68) = 4.917$, $p = .030$, $\eta_p^2 = .067$, which showed that RNG did affect performance relative to the control group (84% vs. 89%). However, none of the interactions were significant (all $ps > .10$).

For the response latencies in Experiment 1, no main effect of instructions emerged, $F(1, 68) = 1.956$, $p = .166$, $\eta_p^2 = .028$, but conflict problems did take longer to complete than no-conflict problems (7,386 vs. 6,797 ms), producing a

Table 2 Experiment 1: Average accuracy and latency scores for belief and logic instructions across both conditions

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No Conflict	Conflict	No Conflict	
RNG					
Average response (%)	70	90	81	93	84
Latency (ms)	9,017	7,761	8,012	8,331	8,280
Control					
Average response (%)	80	93	87	96	89
Latency (ms)	6,575	5,568	5,939	5,526	5,902
Mean accuracy (%) (across each cell)	75	92	84	95	
Mean latency (ms) (across each cell)	7,796	6,665	6,976	6,929	

The results exclude below-chance scores and include correct only latencies. RNG= random number generation ($N=35$), control ($N=35$)

significant main effect of problem type, $F(1, 68)=9.335$, $p=.003$, $\eta_p^2=.121$. We found a main effect of load, $F(2, 68)=20.709$, $p<.001$, $\eta_p^2=.233$, demonstrating a significant increase in latencies when participants were engaged in RNG (8,280 vs. 5,902 ms). There was also an interaction between instruction and problem type, $F(1, 68)=8.340$, $p=.005$, $\eta_p^2=.109$. Follow-up analyses showed that the impact of conflict was significant for belief judgments, $F(1, 68)=19.831$, $p<.001$, $\eta_p^2=.226$, but not for logic judgments, $F(1, 68)=0.028$, $p=.869$, $\eta_p^2<.001$. However, instructions did not interact with load, $F(1, 68)=.093$, $p=.761$, $\eta_p^2=.001$; therefore, load had similar impacts on the latencies for both belief and logic judgments. Finally, none of the remaining interactions were significant (all $ps>.10$).

Discussion

The results from this experiment showed that although participants were more accurate under logic instructions, random number generation had an impact on overall accuracy scores and significantly increased response times. Although participants were less accurate in the RNG group, the secondary load had a similar impact under both instructional sets. These findings are inconsistent with the default interventionist account and suggest that effortful processing is required under both types of instructions.

It has been argued that MP inferences are drawn automatically through direct rules of inference (Braine & O'Brien 1991) that can be triggered by subliminal stimuli and carried out automatically, even without instructions (Reverberi, Burigo, & Cherubini, 2009; Reverberi, Pischedda, Burigo, & Cherubini, 2012; Rips, 1988). However, in Experiment 1, MP arguments were influenced, albeit to a small extent, by a secondary load, which suggests that they are *not* implicit in nature. It could be argued, nevertheless, that the effect of load was the result of RNG interfering with other aspects of

information processing, such as response selection or generation, rather than the activation of a relevant inference rule. In Experiment 2, we introduced a set of more complex arguments in order to examine whether the impact of our secondary task would increase with increased problem complexity.

Experiment 2

Thus far, we have established that RNG affects belief judgments and logic-based MP inferences, suggesting that both judgments depend on Type 2 processing and draw upon executive resources. Furthermore, the findings replicate earlier research in showing that logic-based inferences produce higher accuracy scores than judgments based on belief. One explanation for the higher logic scores could be that with MP arguments, a valid conclusion would be a direct match to the content of the consequent in the major premise. This potentially provides a shortcut for making a validity judgment that would not require integration of the minor premise. Such a shortcut could explain why participants respond more rapidly under logic instructions and are only marginally affected by a secondary task. To rule out this possibility, disjunctive arguments can be used, whereby the conclusion does not directly match a proposition within the major premise. The use of disjunctive arguments would also allow us to determine whether the findings can be extended to these more complex inferences. Reverberi et al. (2012) found that disjunctive arguments did not produce the same priming effects as MP arguments, proposing that disjunctions require more effortful processing. Additionally, Handley et al. (2011) replicated their initial findings using disjunctive arguments. Using more complex arguments, we would expect to find that they are harder to reason with (Johnson-Laird, Byrne, & Schaeken, 1992) and would produce poorer accuracy scores overall. With regard to instruction type, the parallel

competitive account would predict that both belief and logic judgments would be affected by load, whereas the default interventionist account would expect load to have a greater impact on logic-disjunctive arguments under logical instructions.

Our second aim was to examine the extent to which individual differences in reflective processing are associated with performance under different instructions. Individual-difference studies have shown that limitations in working memory are key to explaining reasoning performance (Capon, Handley, & Dennis, 2003). They have established that limitations to an individual's effortful processing can predict logical reasoning (Evans & Over, 1996; Stanovich, 1999). Furthermore, belief-bias studies have demonstrated how different types of reasoning can produce different cognitive demands (Stanovich & West, 2000). With regard to the present study, we are interested in the extents to which belief- and logic-based reasoning are associated with performance on the cognitive reflection task (CRT).

Frederick (2005) introduced the CRT as a measure of an individual's ability to suppress and disregard initial intuitive responses, in order to arrive at the correct, deliberated answer. The three-item questionnaire is said to be a good predictor of the tendency to use heuristics and biases in making judgments and can explain more variance than the typical measures of cognitive ability: executive functioning and thinking dispositions (Toplak, West, & Stanovich, 2011). Toplak et al. described the CRT as a measure that can assess the inclination toward "miserly processing," which corresponds to Type 1 processes in the dual-process literature (De Neys & Glumicic, 2008; Evans, 2008; Kahneman & Frederick, 2002; Sloman, 1996). With this measure, we will look at the relationship between CRT scores, which reflect effortful processing, and judgments made on the basis of the validity of the argument or its believability.

Method

Participants

A total of 80 undergraduates from the University of Plymouth (15 males and 65 females, mean age = 23 years) took part in Experiment 2, in exchange for course credits.

Design and materials

A 2 (Instruction) \times 2 (Problem Type) \times 2 (Complexity) \times 2 (Load) mixed design was used, in which participants were randomly assigned to the RNG (load) or control (no load) group. Each participant received 128 reasoning problems, both MP and disjunctive arguments (Complexity factor) under both belief and logic instructions (Instruction Type), which consisted of both conflict and no-conflict problems (Problem

Type). See Table 3 for more details, and Appendix B for a complete set of the stimuli.

Reasoning task The same 64 MP conditionals from Experiments 1 were used in Experiment 2, with an additional 64 disjunctive arguments. The disjunctives consisted of equal numbers of two types of argument: *denial inferences* (see Disjunctive Type A in Table 3), in which the categorical premise denies one of the propositions in the major premise and the conclusion affirms the remaining one, and *affirmation inferences*, in which the categorical premise affirms one of the propositions in the major premise and the conclusion denies the remaining one (see Disjunctive Type B, Table 3). The use of both denial and affirmation inferences ensured that there would be no confounds between the polarity of the conclusion (i.e., whether it was negative or affirmative) and logical validity.

The disjunctives, like the MPs, consisted of 16 scenarios with two conflict and two no-conflict items within each scenario. As in Experiment 1, counterbalancing the problems required one group of participants to be presented with a conflict and a no-conflict item from each scenario under belief instructions and the remaining two items under logic instructions, and the second group was presented with the same scenarios and items under different instructions. Again, each participant received equal numbers of belief-based and logic-based questions and equal numbers of logically valid and invalid conclusions, along with additional practice trials to include the disjunctive arguments, as displayed in Table 3.

Random number generation We used the same method as in Experiment 1. The average scores for the randomness indices were 3% for redundancy, 60% for RNG, and 40% for adjacency.

Individual difference measures The CRT consisted of the following three questions:

- A bat and a ball cost £1.10 in total. The bat costs £1.00 more than the ball. How much does the ball cost?... ..
- If it takes five machines 5 min to make five widgets, how long would it take 100 machines to make 100 widgets?... ..
- In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake?... .. The CRT was administered in paper format, and participants were required to note down their answers next to each question.

Table 3 Examples of the Disjunctive arguments used in Experiments 2 and 3

Conflict Arguments	No-Conflict Arguments
<p>Disjunctive Type A</p> <p>Argument 1</p> <p>Either the sun is yellow or it is blue.</p> <p>Suppose the sun is not yellow.</p> <p>Does it follow that the sun is blue?</p> <p>Logic: ✓ Belief: X</p> <p>Argument 3</p> <p>Either the sun is yellow or it is blue.</p> <p>Suppose the sun is not yellow.</p> <p>Does it follow that the sun is not blue?</p> <p>Logic: X Belief: ✓</p> <p>Disjunctive Type B</p> <p>Argument 1</p> <p>Either the sea is blue or it is pink.</p> <p>Suppose the sea is pink.</p> <p>Does it follow that the sea is not blue?</p> <p>Logic: ✓ Belief: X</p> <p>Argument 3</p> <p>Either the sea is blue or it is pink.</p> <p>Suppose the sea is pink.</p> <p>Does it follow that the sea is blue?</p> <p>Logic: X Belief: ✓</p>	<p>Disjunctive Type A</p> <p>Argument 2</p> <p>Either the sea is blue or it is pink.</p> <p>Suppose the sea is not pink.</p> <p>Does it follow that the sea is blue?</p> <p>Logic: ✓ Belief: ✓</p> <p>Argument 4</p> <p>Either the sea is blue or it is pink.</p> <p>Suppose the sea is not pink.</p> <p>Does it follow that the sea is not blue?</p> <p>Logic: X Belief: X</p> <p>Disjunctive Type B</p> <p>Argument 2</p> <p>Either the sun is yellow or it is blue.</p> <p>Suppose the sun is yellow.</p> <p>Does it follow that the sun is not blue?</p> <p>Logic: ✓ Belief: ✓</p> <p>Argument 4</p> <p>Either the sun is yellow or it is blue.</p> <p>Suppose the sun is yellow.</p> <p>Does it follow that the sun is blue?</p> <p>Logic: X Belief: X</p>

Procedure

The instructions were the same as in Experiment 1. After participants had completed the reasoning part of the experiment, they were presented with the CRT questions. No time limit was given for completing these questions.

The CRT results were calculated by separating the participants into two groups, the high-CRT group ($n = 29$), which consisted of those that scored 1 or more out of 3, and the low-CRT group ($n = 43$), for those that failed to give a correct answer (see, e.g., Frederick, 2005).

Results

Consistent with Experiment 1, we eliminated those who scored 50% or below on the conflict items for both the RNG and control groups (8 in total). We also removed participants who took longer than 2 s on average to generate random numbers, on the basis that they were not adequately engaged in the secondary task. In total, 13 participants were eliminated, three for exceeding an average of 2 s per number generation, and two for producing stereotyped sequences, which gave a total of $N = 31$ for the RNG group and $N = 36$ for the control group.

A 2 (Instructions: belief/logic) \times 2 (Problem Type: conflict/no conflict) \times 2 (Complexity: MP/disjunctives) \times 2 (Load: RNG/control) mixed design ANOVA was carried out on accuracy measures that were arcsine-transformed. The analysis showed a main effect of instructions, $F(1, 65) = 19.266$, $p < .001$, $\eta_p^2 = .229$, with logic judgments producing the most accurate scores across both groups (87% vs. 81%). A main effect of problem type, $F(1, 65) = 66.225$, $p < .001$, $\eta_p^2 = .505$, reflected poorer performance on conflict than on nonconflict items (78% vs. 90%). There was also a main effect of complexity, $F(1, 65) = 43.964$, $p < .001$, $\eta_p^2 = .403$, in which MP items produced higher accuracy scores than disjunctive arguments (87% vs. 81%), and a main effect of load, $F(2, 65) = 13.193$, $p < .001$, $\eta_p^2 = .169$, demonstrating that performance was poorer under load (80% vs. 88%). However, load did not interact significantly with instructions, $F(1, 65) = 1.968$, $p = .165$, $\eta_p^2 = .029$.

We also observed an interaction between instructions and problem type, $F(1, 65) = 20.278$, $p < .001$, $\eta_p^2 = .239$, showing a bigger impact of conflict on belief problems (72% conflict vs. 90% no-conflict) than on logic problems (84% conflict vs. 90% no conflict). In order to establish whether this interaction was present for both MP and disjunctives, we carried out two separate 2 (Problem Type) \times 2 (Instructions) repeated measures ANOVAs on the two argument types. The results revealed that

Table 4 Experiment 2: Average accuracy and latency scores across the both groups for belief and logic instructions in each problem type, for modus ponens (MP) and disjunctive arguments

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No Conflict	Conflict	No Conflict	
RNG – MP					
Average response (%)	68	91	81	91	83
Latency (ms)	7,747	7,517	7,898	7,376	7,635
Control – MP					
Average response (%)	79	94	90	97	90
Latency (ms)	5,327	5,515	6,398	5,295	5,634
RNG – Disjunctives					
Average response (%)	64	84	77	79	76
Latency (ms)	9,443	9,518	9,652	8,945	9,390
Control – Disjunctives					
Average response (%)	75	89	88	92	86
Latency (ms)	7,431	7,552	6,878	6,781	7,161
Mean accuracy (%) (across each cell)	72	90	84	90	
Mean latency (ms) (across each cell)	7,487	7,526	7,707	7,099	

RNG= random number generation ($N=31$), control ($N=36$)

for MP judgments, instructions and problem type interacted significantly, $F(1, 65)=9.613$, $p=.003$, $\eta_p^2=.129$, with a bigger impact of conflict under belief instruction, which was also confirmed for disjunctive judgments, $F(1, 65)=22.868$, $p<.001$, $\eta_p^2=.260$. Mean accuracy scores can be found in Table 4. Finally, a significant interaction emerged between complexity and problem type, $F(1, 65)=10.962$, $p=.002$, $\eta_p^2=.144$, showing that the difference between conflict and no-conflict problems was larger for MPs than for disjunctives. No other interactions were significant (all $ps>.10$).

Accuracy scores on the individual difference measure were analyzed across both the experimental and control groups, but as a between-subjects factor, by splitting the scorers into high ($N=29$) and low ($N=43$) CRT groups. The five participants removed for poor randomness scores were included in this set of analyses ($N=72$). The results showed no main effect of CRT, $F(2, 70)=0.409$, $p=.524$, $\eta_p^2=.006$, but did produce a significant interaction between instructions and CRT, $F(2, 70)=4.024$, $p=.049$, $\eta_p^2=.054$, showing a bigger difference between the high- and low-CRT groups for logic judgments (90% high vs. 85% low) than for belief judgments (80% high vs. 81% low), and a marginally significant interaction between problem type and CRT, $F(2, 70)=3.235$, $p=.076$, $\eta_p^2=.044$, with less of an effect of conflict in the high-CRT group (81% conflict vs. 90% no conflict) than in the low-CRT group (77% conflict vs. 90% no conflict); see Table 5. The latencies did not produce any significant interactions with CRT.

A mixed-design ANOVA carried out on the latency data showed no main effect of instructions, $F(1, 65)=.808$, $p=$

$.372$, $\eta_p^2=.012$, and no main effect of problem type, $F(1, 65)=2.611$, $p=.111$, $\eta_p^2=.039$. There was, however, a significant main effect of complexity, $F(1, 65)=66.329$, $p<.001$, $\eta_p^2=.505$, and a main effect of load, $F(2, 65)=21.506$, $p<.001$, $\eta_p^2=.049$, confirming that disjunctive arguments take longer to solve than MP arguments and that participants take significantly longer to produce responses when they are engaged in a secondary task. The analysis did not produce any significant interactions for the latency scores.

Discussion

Our aim in Experiment 2 was to look at the effects of a secondary load on disjunctive arguments to determine whether the relatively small impact of load in Experiment 1 would extend to these more complex inferences. Additionally, we were interested in the relationship between the CRT measure and performance under different reasoning instructions. The findings showed a robust effect of the secondary load for both disjunctive and conditional argument types under both sets of instructions. This suggests that both belief and logic judgments require effortful processing. This outcome is in conflict with the current accounts of reasoning, suggesting that belief-based judgments are not available by default and depend upon executive resources. As in Experiment 1, the findings also replicate earlier work, indicating that belief–logic conflict has a greater influence on belief judgments. Taken together, these findings are consistent with a parallel-processing model in

Table 5 Average accuracy scores in Experiment2 for belief and logic instructions for conflict and no-conflict items across high and low Cognitive Reflection Task (CRT) scorers

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No Conflict	Conflict	No Conflict	
High-CRT average response (%)	71	89	90	90	85
Low-CRT average response (%)	73	89	80	90	83
Mean accuracy (%) (across each cell)	72	89	85	90	

The results exclude below chance accuracy scores

which both judgments draw upon executive resources in a Type 2 (belief)/Type 2 (logic) conflict.

Although the effects were small, the CRT results suggest that accuracy scores on belief judgments were virtually the same across the groups, whereas the high-CRT scorers performed significantly better on the logic-based arguments, implying that the cognitive demands associated with the CRT relate to logic-based reasoning more than to belief-based reasoning. This may suggest that distinct executive resources (e.g., inhibition vs. memory updating; see Miyake et al. 2000) are required for belief and logic judgments, an issue that we will return to in the General Discussion.

A characteristic of the experimental task was that participants were often required to switch from trial to trial between responses based upon logic or beliefs. There is good experimental evidence that task switching is demanding of executive resources. One possibility is that RNG impacts performance because of this task-switching requirement. In Experiment3, we manipulated instructions in a blocked design in order to evaluate the influence of RNG under belief and logic instructions when there was no requirement to switch between response types. A blocked design also allowed an evaluation of performance on the secondary task as a function of the primary-task requirements. This is important, given that participants will often sacrifice performance on one task in favor of maintaining performance on the other (Gilhooly et al., 1993; Gilhooly, Logie, & Wynn, 2002; Phillips, 1999). One possibility is that the trade-off between the primary and secondary tasks differs as a function of the judgments being made.

Experiment 3

Task *switching* or *shifting* is an executive function defined by our ability to disengage from an irrelevant task and subsequently to actively engage in a relevant task (Miyake et al., 2000). Others have argued that shifting should be defined as our ability to perform a new operation in the face of proactive interference from having performed a different operation on the same stimuli (Allport & Wylie, 2000). Either way, shifting between mental states is considered an important aspect of executive control required for adequate performance on executive tasks such as

the Wisconsin Card Sorting Task (Miyake et al., 2000). Furthermore, shifting has been shown to sustain considerable temporal costs (Rogers & Monsell, 1995). Consequently, switching between two instructional sets could impact on both accuracy and latency scores. Therefore, in Experiment3 we aimed to control for this possibility, by administering the problems in a blocked design. This would eliminate switching demands and also allow us to interpret performance on the secondary task by analyzing the randomness data for each condition. The design should provide a clearer indication of the impact that RNG has on reasoning performance and allow us to examine the impact of logic and belief judgments on RNG performance.

Method

Participants

The participants were 71 undergraduates from the University of Plymouth (20 male and 51 female, mean age= 20 years) who took part in exchange for course credits.

Design and materials

A 2 (Instructions) \times 2 (Problem Type) \times 2 (Complexity) \times 2 (Load) mixed block design was used in which participants were randomly assigned to the RNG (load) or the control (no load) group. Each participant received the 128 reasoning problems from Experiment2, both MP and disjunctive arguments (Complexity factor) under both belief and logic instructions (Instructions), which consisted of both conflict and no-conflict problems (Problem Type).

Reasoning task A blocked design was used, whereby the 128 problems from Experiment2 were separated into four blocks of 32 arguments. Each block consisted of 16 MP arguments and 16 disjunctives; in two of the blocks, participants received belief instructions only, and in two of the blocks, logic instructions only. In each *belief* block, 16 conclusions were *believable* and 16 were *unbelievable*. In the *logic* blocks, 16 conclusions were *valid* and 16 were *invalid*. Each participant was presented with two belief blocks and two logic blocks, but

participants were randomly allocated to the *belief-first* or *logic-first* order. The content of the blocks alternated, so that the belief-first group were presented with Block 1 under belief instructions, Block 2 under logic instructions, Block 3 under belief instructions, and Block 4 under logic instructions. For the logic-first groups, the reverse order applied. The problems were counterbalanced in the same way as in Experiments 1 and 2.

Random number generation We used the same method as in Experiment 2.

Individual difference measures The CRT, which consisted of the three questions presented in Experiment 2, was administered to each participant in paper format.

Procedure

Before starting the reasoning task, each participant was required to generate a baseline set of random numbers for a total of 5 min. This set of numbers provided a baseline measure of the randomness indices that could be compared against each experimental block. The control group were also asked to provide a baseline measure, but only as a way of ensuring that both groups (RNG and control) started the reasoning questions at the same level of cognitive fatigue following the 5-min task. The 5-min interval was based on previous analysis demonstrating that without a secondary load, participants would take an average of 5 min to complete 32 questions, which was the number of problems allocated to each experimental block.

The instructions were altered to incorporate the baseline measure for each participant. A 5-min timer was built into the computer program, which would automatically instruct the participants to stop counting when the time was up and proceed to the next part of the experiment. The experiment concluded with each participant completing the CRT.

The RGCalc program was used to analyze the randomness of the numbers generated by each participant (see Exp. 1). Since we were measuring randomness and articulation speed, we did not exclude participants for being too slow or for poor randomness scores when analyzing the secondary task. Participants received a score between 0 and 3 for their performance on the CRT.

Results

As in Experiment 2, we eliminated those that scored 50% or below on the conflict items for both the RNG and control groups. A total of ten participants were eliminated, six from the RNG group and four from the control group, which gave us totals of $N=30$ in the RNG group and $N=31$ in the control group.

Randomness data

We looked at the overall randomness indices across the three “primary-task” conditions (belief instructions, logic instructions, and baseline; see Table 6). To compare the three independent variables of randomness, a multivariate ANOVA was carried out to control for Type I errors. The results showed a significant difference in randomness indices for the primary task, $F(6, 208)=5.386$, $p<.001$, $\eta_p^2=.136$. Separate ANOVAs indicated that the primary task had a statistically significant effect on redundancy scores, $F(2, 105)=17.316$, $p<.001$, $\eta_p^2=.248$, but not on adjacency scores, $F(2, 105)=0.458$, $p=.634$, $\eta_p^2=.009$, or RNG scores, $F(2, 105)=0.024$, $p=.976$, $\eta_p^2<.001$. Post-hoc analysis showed that significant differences on redundancy scores occurred between baseline and logic instructions ($p<.001$) and baseline and belief instructions ($p<.001$), but not between belief and logic instructions ($p=.673$). The adjacency and RNG scores were not significantly different across the three primary tasks. Therefore, the randomness data demonstrated that RNG performance was equivalent across the two instructional conditions.

Accuracy data

A mixed-design ANOVA on arcsine-transformed accuracy scores uncovered a main effect of instructions, $F(1, 59)=14.359$, $p<.001$, $\eta_p^2=.196$, confirming that performance was better under logic instructions (88% vs. 82%). We also found a main effect of problem type, $F(1, 59)=40.138$, $p<.001$, $\eta_p^2=.405$, showing poorer performance on conflict items (80% vs. 90%). A main effect of complexity also emerged, $F(1, 59)=22.195$, $p<.001$, $\eta_p^2=.273$, with MP producing higher accuracy rates (87% vs. 83%), and a main effect of load, $F(2, 59)=8.525$, $p=.005$, $\eta_p^2=.126$, with RNG reducing accuracy scores (81% vs. 89%), replicating all of the main effects of Experiment 2. See Table 7 for all accuracy and latency results.

We also observed an interaction between instructions and problem type, $F(2, 59)=9.778$, $p=.003$, $\eta_p^2=.142$, again demonstrating that belief–logic conflict had more of an impact on judgments concerning the believability of the conclusion (74% believability–conflict vs. 90% believability–no

Table 6 Average scores (in percentages) across each measure of randomness for the baseline and experimental conditions in Experiment 3

Variable	Baseline	Belief Instructions	Logic Instructions
Redundancy (%)	1	4	4
RNG (%)	32	34	34
Adjacency (%)	58	57	58

RNG= random number generation

Table 7 Experiment 3: Average accuracy and latency scores, across both groups, for belief and logic instructions in each problem type for modus ponens (MP) and disjunctives (Disj)

Variable	Belief Instructions				Logic Instructions				Overall Means
	Conflict		No Conflict		Conflict		No Conflict		
	MP	Dis	MP	Disj	MP	Disj	MP	Disj	
RNG									
Response accuracy (%)	74	70	87	84	82	77	90	80	81
Latency (ms)	5,460	6,858	5,191	6,743	5,644	7,672	6,408	8,185	6,520
Control									
Response accuracy (%)	78	75	94	92	91	90	97	95	89
Latency (ms)	4,263	5,841	3,962	5,350	4,511	5,775	4,263	5,596	4,945
Mean accuracy (%) (across each cell)	76	73	91	88	87	84	94	88	
Mean latency (ms) (across each cell)	4,861	6,350	4,577	6,047	5,078	6,724	5,336	6,891	

The results exclude below chance scores and include correct only latencies. RNG= random number generation ($N=30$), control ($N=31$)

conflict), rather than its validity (85% logic–conflict vs. 91% logic–no conflict). However, load had a larger impact on logic-based judgments (82% load vs. 93% no load) than on belief-based judgments (79% load vs. 85% no load), as was indicated by an instruction by load interaction, $F(2, 59)=7.295$, $p=.009$, $\eta_p^2=.110$.

Latency data

Analysis carried out on the response latencies produced the following results: a main effect of instructions, $F(1, 59)=5.786$, $p=.019$, $\eta_p^2=.089$, indicating that belief-based judgments took less time to complete (5,458 vs. 6,007 ms), in contrast to Experiments 1 and 2. There was also a main effect of complexity, $F(1, 59)=108.661$, $p<.001$, $\eta_p^2=.648$, with disjunctive arguments taking longer to process than conditional arguments (6,502 vs. 4,962 ms), and a main effect of load, $F(2, 59)=16.596$, $p<.001$, $\eta_p^2=.220$, with longer latencies in the RNG group (6,520 vs. 4,944 ms). We observed no significant interactions (all $ps>.10$).

Individual difference measure

Accuracy data Analysis on the CRT scores produced a marginally significant interaction between instructions, problem type, and CRT, $F(2, 59)=3.774$, $p=.057$, $\eta_p^2=.060$. In order to explore this interaction, both the high ($N=37$) and low ($N=24$) CRT groups were analyzed separately with two 2 (Problem Type) \times 2 (Instructions) \times 2 (Complexity) repeated measures ANOVAs. The results showed an Instructions \times Problem Type interaction for the high-CRT group, $F(1, 36)=14.032$, $p=.001$, $\eta_p^2=.280$, with conflict having a bigger impact on believability than on logic judgments (see

Table 8). The interaction suggested that the belief–logic conflict had more of an impact on believability for those with superior cognitive reflection. In other words, high CRT scorers were more biased by logic under belief instructions. The latencies did not produce any significant interactions with CRT.

Discussion

In Experiment 3, a blocked design was adopted in order to minimize the impact of switching demands on task performance. The idea was to evaluate the impact of RNG on the primary task, independent of the requirement for participants to switch instructional sets as a function of the response options presented. Interestingly, in line with Handley et al. (2011), the impact of conflict was greatest under belief instructions, suggesting that even when participants are aware from trial to trial that they only needed to respond on the basis of the believability of the conclusion, a competing logical structure continued to interfere with their ability to do so. Surprisingly, given the greater difficulty of responding on the basis of beliefs, the impact of RNG was greatest under logic instructions. This did not arise because participants were differentially allocating resources across the primary and secondary tasks. Although redundancy was greater in the random sequences generated under secondary-task conditions, we found no evidence that this differed across instructional conditions. The differential impacts of RNG on logic judgments were reflected in an interaction between load and instructions. This contrasted with Experiments 1 and 2, in which a main effect of load was observed in the absence of an interaction.

It is important to be confident that the absence of an interaction in Experiments 1 and 2 was not the result of a lack of power. Consequently, we completed a post-hoc power

Table 8 Average accuracy scores in Experiment 3 for belief and logic instructions for conflict and no-conflict items across high and low Cognitive Reflection Task (CRT) scorers

Variable	Belief Instructions		Logic Instructions		Overall Means
	Conflict	No Conflict	Conflict	No Conflict	
High-CRT average response (%)	74	92	86	91	86
Low-CRT average response (%)	75	86	83	90	84
Mean accuracy (%) (across each cell)	75	89	85	91	

The results exclude below chance accuracy scores

analysis of Experiments 1 and 2 based on the effect size observed in Experiment 3. This demonstrated that with $N=70$ in Experiment 1 and $N=67$ in Experiment 2, there was adequate power (.8) to detect a medium effect size of the kind observed in Experiment 3, and yet both experiments failed to detect any interaction between instruction and load.

One other important observation in Experiment 3 was that belief judgments were significantly quicker than logic judgments, in contrast to Experiments 1 and 2, in which trials were interspersed. In the General Discussion, we will reflect on these findings while also considering the implications of the CRT findings for understanding the processes underpinning belief and logic judgments.

General discussion

The objective of the experiments presented here was to evaluate the degree to which logic and beliefs interfere with one another when participants were engaged in a secondary task that burdened working memory resources. Previous research has shown, contrary to the default interventionist dual-process account, that participants made more errors and generally took longer to judge the believability of a conclusion when this conclusion conflicted with what was logically entailed. The three experiments in this article have also demonstrated that belief-based judgments on conflict items produce more errors than logic-based judgments and that conflict between belief and logic results in a greater impact on judging the believability of a conclusion.

Experiments 1 and 2 showed that RNG affects *both* belief and logic judgments, suggesting that both depend on effortful processing. Experiment 3 employed a blocked design to eliminate possible switching demands on the reasoning task. Interestingly, the principal findings regarding accuracy were replicated in the blocked design, but RNG only impacted significantly on the logic judgments. Importantly, the impact of RNG was not linked to the resolution of conflict, since there was no evidence of greater effects of the secondary task on conflict than on no-conflict items. This suggests that RNG interfered with the structural processing underpinning logical judgments rather than

with the resolution of conflict between competing responses. It is worth noting that, despite an effect of secondary load, logical accuracy on these problems remained high. Given the relative simplicity of the arguments used, this is perhaps unsurprising; however, it does suggest that even simple conditional and disjunctive inferences require some cognitive effort.

A common argument in the dual-process literature is that logical reasoning is effortful and belief judgments are less so; however, the behavioral data we have presented suggest otherwise. How might we reconcile the accuracy data, which suggest that belief judgments are most demanding, with the secondary-task data, which show greatest interference with logic judgments? In our view, the data are consistent with a Type 2/Type 2 conflict that arises because of competing responses. We suggest that the logical response is available early, but nevertheless requires participants to effortfully reason from the underlying structure of the presented argument. This claim is consistent with research that has linked reasoning from simple propositional arguments with measures of cognitive ability (see, e.g., Newstead, Handley, Harley, Wright, & Farrelly, 2004). In contrast, the belief-based response is available later, and successful responding depends on the inhibition of an available and competing response. The critical feature of this account is that both types of judgments depend on Type 2 processing, but rely on distinct executive resources. This account is consistent with a newly developed parallel competitive model (Handley et al., 2011; Handley & Trippas, 2015), which suggests that both belief- and logic-based judgments are activated early but, in order to answer according to beliefs, a readily available logical response requires inhibition. This would explain why conflict between belief and logic consistently results in a greater impact on judgments of conclusion believability.

However, it is important to note that belief judgments were not consistently slower than logic judgments; in fact, the pattern reversed in Experiment 3 when problems were presented in a blocked design. Despite this, the greater effect of conflict under belief instructions remained. Interestingly, this pattern was also observed in Experiments 1 and 2 of Handley et al. (2011), in which participants were also presented problems in a blocked design or with an instructional cue prior to item

presentation. Handley et al. (2011) argued that these design features allow the development of a shortcut strategy on belief items, whereby a response could be generated by simply referring to the categorical premise and conclusion (on MP items) or the conclusion alone (on disjunctive arguments). If some participants adopted this strategy, it would explain the reduced latencies in Experiment 3.

How, then, do we reconcile our findings with traditional belief-bias studies that have shown belief bias increasing under time pressure (Evans & Curtis-Holmes, 2005) or under cognitive load (De Neys, 2006)? We argue that beliefs do not necessarily function through the activation of a Type 1, default process. The degree of interference of beliefs with reasoning will depend upon the complexity of the logical reasoning required (see Handley & Trippas, 2015, for a detailed model). The greater the complexity of the logical reasoning, the longer the process will take, and the greater the opportunity for beliefs to interfere with reasoning. Hence, the belief bias observed on complex syllogisms arises because the logical processes take longer to complete, so that a belief-based response is available prior to a logical one. On simpler logical problems, when participants are instructed to reason on the basis of belief, the fast-completing logical response interferes with a belief-based output.

A key characteristic of the RNG task is that the successful generation of random numbers depends on minimizing structure in the sequence of responses generated. The task also requires updating of working memory to ensure that generated sequences do not repeat previous ones. It is our conjecture that RNG consequently interferes with the ability to extract the underlying structure of an argument required for processing the validity of an inference, but does not impact upon the capacity to inhibit the logical response—hence, the larger effect of RNG on logic-based judgments in Experiment 3. Given this explanation, an interesting topic for future research might focus on examining the impact of different types of secondary tasks on performance. For example, we might expect less interference from a task that places minimal demands on structural processing (e.g., the dot memory task; De Neys, 2006).

In order to explain the shift from the main effect of load (Exps. 1 and 2) to its increased impact on logic judgments, we have assumed that the switching between instructional sets (disengaging from one irrelevant task and effectively engaging in a relevant one) places additional demands on executive resources. These switching demands combined with RNG increase the executive demands of the task consequently impacting upon both types of judgment.

As we argued above, this account implicates effortful processing in both judgments types and a distinction in the executive resources required for each type of judgment. This proposal receives some support from the results of the CRT (Frederick, 2005). In Experiment 2, participants with high CRT scores show higher accuracy on logical judgments. Similarly, in Experiment 3 high CRT is associated with a greater impact of conflict on belief

judgments. Both effects suggest that higher CRT scores are associated with greater sensitivity to logical structure. Interestingly, the data do not show that the CRT is linked to less belief bias under logical instructions. This suggests that successful performance on the CRT relates to a person's ability to extract the core structure of the problem in order to make the correct judgment. The evidence of a relationship between high CRT and logical judgments supports the idea that CRT reflects sensitivity toward the underlying structure of an argument.

Furthermore, we found some evidence to suggest that that high CRT scorers were less random on the secondary task (i.e., higher RNG scores) as their performance under logical instructions increased, $r(19) = .652, p = .002$, which implies that structural processing impaired their ability to generate random response alternatives. In contrast, the low-CRT group showed the opposite trend, $r(35) = -.511, p = .043$. Therefore, the results suggest that those with higher reflective ability are not more proficient at overcoming conflict but are better at extracting the underlying structure of a problem, which interferes with their ability to abandon structural processing on the secondary task.

Could there be an alternative explanation for our findings? Is it possible, for example, that explicitly instructing participants to make judgments on the basis of belief changes the nature of the processes involved, perhaps leading participants to engage in an explicit Type 2 process of belief evaluation? Although we cannot fully discount this possibility, it is important to note that belief judgments are *more* prone to errors and generally take *longer* than logic judgments. Even though participants are explicitly instructed to give belief responses, there is still no basis to expect these judgments to result in more errors and to be subject to greater interference from conflict than are logical judgments. At the very least, the findings show that explicit belief judgments are subject to logical interference, which in itself is a novel finding. One way of determining whether the same processes underlie explicit belief judgments and “belief-biased” responses under logical instructions is to examine the latencies associated with each type of response. If different processes underpin each judgment then one would expect the former to take longer than the latter. We reanalyzed latencies for conflict problems from Experiment 3, in which the blocked design allowed a cleaner evaluation of latencies for correct and incorrect responses. Surprisingly, the latencies associated with incorrect “belief-biased” responses to conflict problems under logical instructions ($M = 6,249$ ms) were, if anything, slightly longer than the latencies associated with correct belief-based responses under belief instructions ($M = 5,707$ ms), although this difference was not reliable, $t(47) = 1.07, p = .29$.¹ The absence of a

¹ Note that the degrees of freedom for this comparison reflect a reduced sample size, given that a number of participants did not make any errors to logic conflict problems, and hence they were not included in the comparison.

difference in latencies as a function of response category is inconsistent with an explanation of the findings based on a switch to deliberative belief-based reasoning under instruction.

Is it possible instead that the difficulty of belief conflict judgments arises because of the requirement to regularly switch from logic- to belief-based trials? Perhaps the inhibition of beliefs on logic trials has negative consequences for subsequent trials in which beliefs have to be activated? Fortunately, the blocked design of Experiment 3 allowed a direct test of this possibility, through an analysis of the first two blocks. Recall that participants received the problem blocks in counterbalanced order, so that half of the participants received a logic block followed by a belief block, and the remaining group received the same problems in reverse order. If the completion of logic trials interferes with subsequent belief trials, we would expect the second block of belief judgments to be more difficult than the first. We reanalyzed the first two blocks of Experiment 3, including order (belief items first or logic items first) as an additional variable. Importantly, the analysis revealed an interaction between problem type and instructions, $F(1, 57) = 7.44$, $p < .01$, $\eta_p^2 = .115$, once again showing that conflict had more of an impact on belief judgments (75% believability–conflict vs. 90% believability–no conflict) than logic judgments (87% logic–conflict vs. 90% logic–no conflict). Of key relevance here, we also found an interaction between instructions and order, $F(1, 57) = 10.77$, $p < .01$, $\eta_p^2 = .19$. However, there was no evidence that completing a logic block first led to poorer performance on a subsequent block of belief problems. In fact, performance on the second block of belief problems was, if anything, slightly improved after a block of logic problems ($M = 86\%$), as compared to the first block of belief problems ($M = 80\%$), although this difference was not significant, $F(2, 57) = 1.769$, $p = .189$, $\eta_p^2 = .030$. We observed a similar improvement in performance in logical reasoning, with the second block of logic problems showing higher accuracy after a block of belief problems ($M = 92\%$), as compared to the first block of logic problems ($M = 85\%$). In this case, the difference was reliable, $F(2, 57) = 4.278$, $p = .043$, $\eta_p^2 = .070$. This analysis principally shows a practice effect, whereby participants are more accurate on the second block of problems, irrespective of the block they received first. There is no evidence that inhibiting beliefs on a block of logic problems subsequently leads to poorer performance in judging believability.

Finally, we consider whether the believability of the problem premises might explain some of our findings. A characteristic of the conflict problems is that the premises are always unbelievable. Perhaps making judgments from unbelievable premises results in more deliberative processing, and consequently explains the impact of secondary load on performance? Once again, our findings are inconsistent with this suggestion; load did not differentially impact problem type,

signifying that load does not uniquely interfere with conflict problems that have unbelievable premises. Is it possible, instead, that participants misunderstood the belief instructions or were confused by them, and consequently performed poorly on these items? Once again, our data suggest not. Participants showed excellent levels of accuracy on no-conflict belief items, suggesting that they were clearly able to evaluate the believability of the presented conclusions in response to the cue provided. Although a number of participants were removed from the analysis because their accuracy on conflict items was at chance level, a significant proportion of participants across the experiments were eliminated because of below-chance performance on logic conflict items (36%). Although a greater proportion (54%) showed chance or below-chance performance on belief judgments, this simply reflects the fact that belief conflict items were more difficult to solve than their logical counterparts.

It is important to note that our account does not necessitate abandoning a dual-process approach to reasoning. We agree that often there can be competing cues to a solution and that this can lead to error. Where our account differs is that we assume that both belief- and logic-based judgments depend on Type 1 and Type 2 processes and that interference can as readily arise between logic and beliefs as the other way around. Some researchers have, in contrast, argued that the evidence used to support separate systems is more consistent with a single-system account (see, e.g., Kruglanski & Gigerenzer, 2011; Osman, 2004). Keren and Schul (2009) proposed a uni-model and argued that the presence of conflict is not adequate evidence for two independent systems. They suggest that the belief-bias effect can be explained by assuming that logical validity and believability are two distinct types of external criteria that a single system can use to evaluate a conclusion, and that the level or awareness of alternative responses fluctuates continuously when one is reasoning. In other words, a single system can shift between many mental states to solve different tasks, and these states are defined by an assembly of different features such as speed, level of control, and awareness. In some respects, this account is not a bad fit to the data presented here, and it has some elements in common with the parallel-processing model, such as the proposal that conflict arises through the activation of competing cues to the solution. However, we argue that there are advantages in retaining the distinction between Type 1 and Type 2 processes, since inevitably, certain judgments will require greater reliance on explicit processing than others.

Conclusion

The overall aim of this article was to establish whether engaging in a secondary task while reasoning on the basis of belief or logic would have differential impacts on performance. Our findings suggest that responding under both instruction types

requires some effortful processing, which is inconsistent with prototypical dual-process accounts of belief bias. The fact that logic has a greater impact on judging the believability of a conclusion suggests that logical inferences can be accomplished rapidly, hence interfering with belief-based

judgments. The impact of a complex executive load on simple logical judgments suggests that even these draw on cognitive resources. A task for future research will be to map the distinct executive demands of reasoning on the basis of beliefs or logical structure.

Appendix A: Experiments 1–3, modus ponens stimuli

The full 64 conditional arguments are listed below. In each experiment, one group received half under logic instructions and the other half under belief instructions, and these

assignments were counterbalanced against the second group in each experiment.

Experimental trial stimuli

Premise 1	Premise 2	Conclusion
If a child is crying, then it is happy.	Suppose a child is crying.	Does it follow that the child is sad?
If a child is crying, then it is happy.	Suppose a child is crying.	Does it follow that the child is happy?
If a child is crying, then it is sad.	Suppose a child is crying.	Does it follow that the child is happy?
If a child is crying, then it is sad.	Suppose a child is crying.	Does it follow that the child is sad?
If a dog is barking, then it is silent.	Suppose a dog is barking.	Does it follow that the dog is silent?
If a dog is barking, then it is silent.	Suppose a dog is barking.	Does it follow that the dog is loud?
If a dog is barking, then it is loud.	Suppose a dog is barking.	Does it follow that the dog is loud?
If a dog is barking, then it is loud.	Suppose a dog is barking.	Does it follow that the dog is silent?
If a person is swimming, then he is dry.	Suppose a person is swimming.	Does it follow that the person is wet?
If a person is swimming, then he is dry.	Suppose a person is swimming.	Does it follow that the person is dry?
If a person is swimming, then he is wet.	Suppose a person is swimming.	Does it follow that the person is dry?
If a person is swimming, then he is wet.	Suppose a person is swimming.	Does it follow that the person is wet?
If the sky is grey, then it is sunny.	Suppose the sky is grey.	Does it follow that it is cloudy?
If the sky is grey, then it is sunny.	Suppose the sky is grey.	Does it follow that it is sunny?
If the sky is grey, then it is cloudy.	Suppose the sky is grey.	Does it follow that it is sunny?
If the sky is grey, then it is cloudy.	Suppose the sky is grey.	Does it follow that it is cloudy?
If a new computer is high-end, then it is cheap.	Suppose a new computer is high-end.	Does it follow that the new computer is expensive?
If a new computer is high-end, then it is cheap.	Suppose a new computer is high-end.	Does it follow that the new computer is cheap?
If a computer is high-end, then it is expensive.	Suppose a new computer is high-end.	Does it follow that the new computer is expensive?
If a computer is high-end, then it is expensive.	Suppose a new computer is high-end.	Does it follow that the new computer is cheap?
If a person eats too much, then they are skinny.	Suppose a person eats too much.	Does it follow that the person is skinny?
If a person eats too much, then they are skinny.	Suppose a person eats too much.	Does it follow that the person is fat?
If a person eats too much, then they are fat.	Suppose a person eats too much.	Does it follow that the person is skinny?
If a person eats too much, then they are fat.	Suppose a person eats too much.	Does it follow that the person is fat?
If a glass falls on the floor then it is intact.	Suppose a glass falls on the floor.	Does it follow that the glass is intact?
If a glass falls on the floor then it is intact.	Suppose a glass falls on the floor.	Does it follow that the glass is broken?
If a glass falls on the floor then it is broken.	Suppose a glass falls on the floor.	Does it follow that the glass is intact?
If a glass falls on the floor then it is broken.	Suppose a glass falls on the floor.	Does it follow that the glass is broken?
If the light switch is on, then it is dark inside.	Suppose the light switch is on.	Does it follow that it is dark inside?

(continued)

If the light switch is on, then it is dark inside.	Suppose the light switch is on.	Does it follow that it is bright inside?
If the light switch is on, then it is bright inside.	Suppose the light switch is on.	Does it follow that it is bright inside?
If the light switch is on, then it is bright inside.	Suppose the light switch is on.	Does it follow that it is dark inside?
If a singer has a sore throat, his singing will be nice.	Suppose the singer has a sore throat.	Does it follow that his singing will be nice?
If a singer has a sore throat, his singing will be nice.	Suppose the singer has a sore throat.	Does it follow that his singing will be bad?
If a singer has a sore throat, his singing will be bad.	Suppose the singer has a sore throat.	Does it follow that his singing will be bad?
If a singer has a sore throat, his singing will be bad.	Suppose the singer has a sore throat.	Does it follow that his singing will be nice?
If an aeroplane runs out of fuel, then it will fly.	Suppose an aeroplane runs out of fuel.	Does it follow that the aeroplane will fly?
If an aeroplane runs out of fuel, then it will fly.	Suppose an aeroplane runs out of fuel.	Does it follow that the aeroplane will crash?
If an aeroplane runs out of fuel, then it will crash.	Suppose an aeroplane runs out of fuel.	Does it follow that the aeroplane will fly?
If an aeroplane runs out of fuel, then it will crash.	Suppose an aeroplane runs out of fuel.	Does it follow that the aeroplane will crash?
If heat bill has not been paid in time then it is warm inside.	Suppose the heat bill has not been paid in time.	Does it follow that it is cold inside?
If heat bill has not been paid in time then it is warm inside.	Suppose the heat bill has not been paid in time.	Does it follow that it is warm inside?
If the heat bill has not been paid in time then it is cold inside.	Suppose the heat bill has not been paid in time.	Does it follow that it is warm inside?
If heat bill has not been paid in time then it is cold inside.	Suppose the heat bill has not been paid in time.	Does it follow that it is cold inside?
If a bear catches a fish, it is slow.	Suppose a bear catches a fish.	Does it follow that the bear is slow?
If a bear catches a fish, it is slow.	Suppose a bear catches a fish.	Does it follow that the bear is quick?
If a bear catches a fish, it is quick.	Suppose a bear catches a fish.	Does it follow that the bear is quick?
If a bear catches a fish, it is quick.	Suppose a bear catches a fish.	Does it follow that the bear is slow?
If a toilet has been flushed, then it is dirty.	Suppose a toilet is flushed.	Does it follow that the toilet is clean?
If a toilet has been flushed, then it is dirty.	Suppose a toilet is flushed.	Does it follow that the toilet is dirty?
If a toilet has been flushed, then it is clean.	Suppose a toilet is flushed.	Does it follow that the toilet is clean?
If a toilet has been flushed, then it is clean.	Suppose a toilet is flushed.	Does it follow that the toilet is dirty?
If a tree is old, then it is small.	Suppose a tree is old.	Does it follow that the tree is small?
If a tree is old, then it is small.	Suppose a tree is old.	Does it follow that the tree is large?
If a tree is old, then it is large.	Suppose a tree is old.	Does it follow that the tree is large?
If a tree is old, then it is large.	Suppose a tree is old.	Does it follow that the tree is small?
If your mother bakes cookies, you will be sad.	Suppose your mother bakes cookies.	Does it follow that you will be happy?
If your mother bakes cookies, you will be sad.	Suppose your mother bakes cookies.	Does it follow that you will be sad?
If your mother bakes cookies, you will be happy.	Suppose your mother bakes cookies.	Does it follow that you will be sad?
If your mother bakes cookies, you will be happy.	Suppose your mother bakes cookies.	Does it follow that you will be happy?
If a hamster gets fed, then it will die.	Suppose a hamster gets fed.	Does it follow that the hamster will live?
If a hamster gets fed, then it will die.	Suppose a hamster gets fed.	Does it follow that the hamster will die?
If a hamster gets fed, then it will live.	Suppose a hamster gets fed.	Does it follow that the hamster will live?
If a hamster gets fed, then it will live.	Suppose a hamster gets fed.	Does it follow that the hamster will die?

Appendix B: Experiments 2–3, disjunctive stimuli

The full 64 disjunctive arguments are listed below. In each experiment, one group received half under logic instructions and the other half under belief instructions, and these

assignments were counterbalanced against the second group in each experiment.

Experimental trial stimuli

Premise 1	Premise 2	Conclusion	Item Type
Either cats are mammals or they are plants.	Suppose a cat is not a mammal.	Does it follow that the cat is not a plant?	AFF
Either cats are mammals or they are plants.	Suppose a cat is not a mammal.	Does it follow that the cat is a plant?	AFF
Either cats are mammals or they are plants.	Suppose a cat is not a plant.	Does it follow that the cat is not a mammal?	AFF
Either cats are mammals or they are plants.	Suppose a cat is not a plant.	Does it follow that the cat is a mammal?	AFF
Either the sky is blue or it is green.	Suppose the sky is not blue.	Does it follow that the sky is green?	AFF

(continued)

Either the sky is blue or it is green.	Suppose the sky is not blue.	Does it follow that the sky is not green?	AFF
Either the sky is blue or it is green.	Suppose the sky is not green.	Does it follow that the sky is blue?	AFF
Either the sky is blue or it is green.	Suppose the sky is not green.	Does it follow that the sky is not blue?	AFF
Either the sun is yellow or it is blue.	Suppose the sun is blue.	Does it follow that the sun is yellow?	DEN
Either the sun is yellow or it is blue.	Suppose the sun is blue.	Does it follow that the sun is not yellow?	DEN
Either the sun is yellow or it is blue.	Suppose the sun is yellow.	Does it follow that the sun is blue?	DEN
Either the sun is yellow or it is blue.	Suppose the sun is yellow.	Does it follow that the sun is not blue?	DEN
Either the sea is blue or it is pink.	Suppose the sea is pink.	Does it follow that the sea is not blue?	DEN
Either the sea is blue or it is pink.	Suppose the sea is pink.	Does it follow that the sea is blue?	DEN
Either the sea is blue or it is pink.	Suppose the sea is blue.	Does it follow that the sea is not pink?	DEN
Either the sea is blue or it is pink.	Suppose the sea is blue.	Does it follow that the sea is pink?	DEN
Either parrots can fly or they can swim.	Suppose a parrot cannot fly.	Does it follow that the parrot cannot swim?	AFF
Either parrots can fly or they can swim.	Suppose a parrot cannot fly.	Does it follow that the parrot can swim?	AFF
Either parrots can fly or they can swim.	Suppose a parrot cannot swim.	Does it follow that the parrot cannot fly?	AFF
Either parrots can fly or they can swim.	Suppose a parrot cannot swim.	Does it follow that the parrot can fly?	AFF
Either obese people are fat or they are skinny.	Suppose an obese person is not fat.	Does it follow that the obese person is skinny?	AFF
Either obese people are fat or they are skinny.	Suppose an obese person is not fat.	Does it follow that the obese person is not skinny?	AFF
Either obese people are fat or they are skinny.	Suppose an obese person is not skinny.	Does it follow that the obese person is fat?	AFF
Either obese people are fat or they are skinny.	Suppose an obese person is not skinny.	Does it follow that the obese person is not fat?	AFF
Either skyscrapers are huge or they are tiny.	Suppose a skyscraper is not huge.	Does it follow that the skyscraper is tiny?	DEN
Either skyscrapers are huge or they are tiny.	Suppose a skyscraper is not huge.	Does it follow that the skyscraper is not tiny?	DEN
Either skyscrapers are huge or they are tiny.	Suppose a skyscraper is not tiny.	Does it follow that the skyscraper is not huge?	DEN
Either skyscrapers are huge or they are tiny.	Suppose a skyscraper is not tiny.	Does it follow that the skyscraper is huge?	DEN
Either roses are flowers or they are machines.	Suppose a rose is a machine.	Does it follow that the rose is not a flower?	DEN
Either roses are flowers or they are machines.	Suppose a rose is a machine.	Does it follow that the rose is a flower?	DEN
Either roses are flowers or they are machines.	Suppose a rose is a flower.	Does it follow that the rose is not a machine?	DEN
Either roses are flowers or they are machines.	Suppose a rose is a flower.	Does it follow that the rose is a machine?	DEN
Either sentences are made out of words or they are made out of bricks.	Suppose a sentence is not made out of words.	Does it follow that the sentence is not made out of bricks?	AFF
Either sentences are made out of words or they are made out of bricks.	Suppose a sentence is not made out of words.	Does it follow that the sentence is made out of bricks?	AFF
Either sentences are made out of words or they are made out of bricks.	Suppose a sentence is not made out of bricks.	Does it follow that the sentence is not made out of words?	AFF
Either sentences are made out of words or they are made out of bricks.	Suppose a sentence is not made out of bricks.	Does it follow that the sentence is made out of words?	AFF
Either mice eat cheese or they eat steel.	Suppose a mouse does not eat cheese.	Does it follow that the mouse eats steel?	AFF
Either mice eat cheese or they eat steel.	Suppose a mouse does not eat cheese.	Does it follow that the mouse does not eat steel?	AFF
Either mice eat cheese or they eat steel.	Suppose a mouse does not eat steel.	Does it follow that the mouse eats cheese?	AFF
Either mice eat cheese or they eat steel.	Suppose a mouse does not eat steel.	Does it follow that the mouse does not eat cheese?	AFF
Either alligators eat meat or they eat plastic.	Suppose an alligator eats plastic.	Does it follow that the alligator does not eat meat?	DEN
Either alligators eat meat or they eat plastic.	Suppose an alligator eats plastic.	Does it follow that the alligator eats meat?	DEN
Either alligators eat meat or they eat plastic.	Suppose an alligator eats meat.	Does it follow that the alligator does not eat plastic?	DEN
Either alligators eat meat or they eat plastic.	Suppose an alligator eats meat.	Does it follow that the alligator eats plastic?	DEN
Either circles are round or they are square.	Suppose a circle is square.	Does it follow that the circle is round?	DEN
Either circles are round or they are square.	Suppose a circle is square.	Does it follow that the circle not round?	DEN
Either circles are round or they are square.	Suppose a circle is round.	Does it follow that the circle is square?	DEN
Either circles are round or they are square.	Suppose a circle is round.	Does it follow that the circle is not square?	DEN
Either monkeys are primates or they are rodents.	Suppose a monkey is not a primate.	Does it follow that the monkey is not a rodent?	AFF
Either monkeys are primates or they are rodents.	Suppose a monkey is not a primate.	Does it follow that the monkey is a rodent?	AFF
Either monkeys are primates or they are rodents.	Suppose a monkey is not a rodent.	Does it follow that the monkey is not a primate?	AFF
Either monkeys are primates or they are rodents.	Suppose a monkey is not a rodent.	Does it follow that the monkey is a primate?	AFF
Either flowers are organic or they are vehicles.	Suppose a flower is not organic.	Does it follow that the flower is a vehicle?	AFF
Either flowers are organic or they are vehicles.	Suppose a flower is not organic.	Does it follow that the flower is not a vehicle?	AFF
Either flowers are organic or they are vehicles.	Suppose a flower is not a vehicle.	Does it follow that the flower is organic?	AFF
Either flowers are organic or they are vehicles.	Suppose a flower is not a vehicle.	Does it follow that the flower is not organic?	AFF
Either shovels are tools or they are stationary.	Suppose a shovel is stationary.	Does it follow that the shovel is not a tool?	DEN

(continued)

Either shovels are tools or they are stationary.	Suppose a shovel is stationary.	Does it follow that the shovel is a tool?	DEN
Either shovels are tools or they are stationary.	Suppose a shovel is a tool.	Does it follow that the shovel is stationary?	DEN
Either shovels are tools or they are stationary.	Suppose a shovel is a tool.	Does it follow that the shovel is not stationary?	DEN
Either spears are weapons or they are thermostats.	Suppose a spear is a thermostat.	Does it follow that the spear is a weapon?	DEN
Either spears are weapons or they are thermostats.	Suppose a spear is a thermostat.	Does it follow that the spear is not a weapon?	DEN
Either spears are weapons or they are thermostats.	Suppose a spear is a weapon.	Does it follow that the spear is not a thermostat?	DEN
Either spears are weapons or they are thermostats.	Suppose a spear is a weapon.	Does it follow that the spear is a thermostat?	DEN

Item Type refers to the category of disjunctives used. AFFare disjunctives of the following type: A or B, not A, therefore B. DEN are of a different type: A or B, A, therefore not B. These are valid because they are exclusive disjunctives, meaning that A and B can never be true at the same time.

References

- Allport, A., & Wylie, G. (2000). Task switching, stimulus–response bindings, and negative priming. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and performance XVIII* (pp. 35–70). Cambridge, MA: MIT Press.
- Baddeley, A. (1986). *Working memory*. Oxford, UK: Oxford University Press, Clarendon Press.
- Braine, M. D., & O'Brien, D. P. (1991). A theory of if: A lexical entry, reasoning program, and pragmatic principles. *Psychological Review*, 98, 182–203. doi:10.1037/0033-295X.98.2.182
- Brown, R., Soliveri, P., & Jahanshahi, M. (1998). Executive processes in Parkinsons disease—Random Number generation and response suppression. *Neuropsychologia*, 36, 1355–1362.
- Capon, A., Handley, S., & Dennis, I. (2003). Working memory and reasoning: An individual differences perspective. *Thinking & Reasoning*, 9, 203–244.
- Chaiken, S., & Trope, Y. (1999). *Dual-process theories in social psychology*. New York, NY: Guilford Press.
- De Neys, W. (2006). Dual processing in reasoning two systems but one reasoner. *Psychological Science*, 17, 428–433.
- De Neys, W. (2012). Bias and conflict: A case for logical intuitions. *Perspectives on Psychological Science*, 7, 28–38. doi:10.1177/1745691611429354
- De Neys, W., & Glumicic, T. (2008). Conflict monitoring in dual process theories of thinking. *Cognition*, 106, 1248–1299.
- Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, 7, 454–459. doi:10.1016/j.tics.2003.08.012
- Evans, J. S. B. T. (2006). The heuristic-analytic theory of reasoning: Extension and evaluation. *Psychonomic Bulletin & Review*, 13(3), 378–395.
- Evans, J. S. B. T. (2007). On the resolution of conflict in dual process theories of reasoning. *Thinking & Reasoning*, 13, 321–339.
- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59, 255–278. doi:10.1146/annurev.psych.59.103006.093629
- Evans, J. S. B. T. (2009). How many dual-process theories do we need? One, two, or many? In J. S. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 33–54). Oxford, UK: Oxford University Press. doi:10.1093/acprof:oso/9780199230167.003.0002
- Evans, J. S. B. T. (2011). Dual-process theories of reasoning: Contemporary issues and developmental applications. *Developmental Review*, 31(2–3), 86–102. doi:10.1016/j.dr.2011.07.007
- Evans, J. S. B. T., Barston, J. L., & Pollard, P. (1983). On the conflict between logic and belief in syllogistic reasoning. *Memory & Cognition*, 11, 295–306.
- Evans, J. S. B. T., & Curtis-Holmes, J. (2005). Rapid responding increases belief bias: Evidence for the dual-process theory of reasoning. *Thinking & Reasoning*, 11, 382–389.
- Evans, J. S. B. T., Handley, S. J., & Bacon, A. M. (2009). Reasoning under time pressure. *Experimental Psychology*, 56, 77–83. doi:10.1027/1618-3169.56.2.77
- Evans, J. S. B. T., & Over, D. E. (1996). *Rationality and reasoning*. Hove, UK: Psychology Press.
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition advancing the debate. *Perspectives on Psychological Science*, 8, 223–241. doi:10.1177/1745691612460685
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19, 25–42. doi:10.1257/089533005775196732
- Gilhooly, K. J., Logie, R. H., Wetherick, N. E., & Wynn, V. (1993). Working memory and strategies in syllogistic-reasoning tasks. *Memory & Cognition*, 21, 115–124. doi:10.3758/BF03211170
- Gilhooly, K. J., Logie, R. H., & Wynn, V. (2002). Syllogistic reasoning tasks and working memory: Evidence from sequential presentation of premises. *Current Psychology*, 21, 111–120.
- Handley, S. J., Capon, A., Beveridge, M., Dennis, I., & Evans, J. S. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking & Reasoning*, 10, 175–195. doi:10.1080/13546780442000051
- Handley, S. J., Newstead, S. E., & Trippas, D. (2011). Logic, beliefs, and instruction: A test of the default interventionist account of belief bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 37, 28–43. doi:10.1037/a0021098
- Handley, S. J., & Trippas, D. (2015). Dual processes and the interplay between knowledge and structure: A new parallel processing model. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 62, pp. 33–58). San Diego, CA: Elsevier Academic Press.

- Horne, R. L., Evans, F. J., & Orme, M. T. (1982). Random number generation, psychopathology, and therapeutic change. *Archives of General Psychiatry*, 39, 680–683.
- Johnson-Laird, P. N., Byrne, R. M., & Schaeken, W. (1992). Propositional reasoning by model. *Psychological Review*, 99, 418–439. doi:10.1037/0033-295X.99.3.418
- Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). Cambridge, UK: Cambridge University Press.
- Keren, G., & Schul, Y. (2009). Two is not always better than one: A critical evaluation of two-system theories. *Perspectives on Psychological Science*, 4, 533–550. doi:10.1111/j.1745-6924.2009.01164.x
- Kruglanski, A. W., & Gigerenzer, G. (2011). Intuitive and deliberate judgments are on the basis of common principles. *Psychological Review*, 118, 97–109. doi:10.1037/a0020762
- Leo, P. D., & Greene, A. J. (2008). Is awareness necessary for true inference? *Memory & Cognition*, 36, 1079–1086. doi:10.3758/MC.36.6.1079
- Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex “frontal lobe” tasks: A latent variable analysis. *Cognitive Psychology*, 41, 49–100. doi:10.1006/cogp.1999.0734
- Morsanyi, K., & Handley, S. J. (2008). How smart do you need to be to get it wrong? The role of cognitive capacity in the development of heuristic-based judgment. *Journal of Experimental Child Psychology*, 99, 18–36.
- Newstead, S. E., Handley, S. J., Harley, C., Wright, H., & Farrelly, D. (2004). Individual differences in deductive reasoning. *Quarterly Journal of Experimental Psychology*, 57A, 33–60. doi:10.1080/02724980343000116
- Osman, M. (2004). An evaluation of dual-process theories of reasoning. *Psychonomic Bulletin & Review*, 11, 988–1010. doi:10.3758/BF03196730
- Phillips, L. H. (1999). The role of memory in the Tower of London task. *Memory*, 7, 209–231.
- Rader, A. W., & Sloutsky, V. M. (2002). Processing of logically valid and logically invalid conditional inferences in discourse comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 59–68. doi:10.1037/0278-7393.28.1.59
- Reber, A. S. (1996). *Implicit learning and tacit knowledge: An essay on the cognitive unconscious*. New York, NY: Oxford University Press.
- Reverberi, C., Burigo, M., & Cherubini, P. (2009). A subliminal premise can automatically trigger an elementary deductive inference. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the 31st Annual Conference of the Cognitive Science Society* (pp. 201–205). Austin, TX: Cognitive Science Society.
- Reverberi, C., Pischella, D., Burigo, M., & Cherubini, P. (2012). Deduction without awareness. *Acta Psychologica*, 139, 244–253.
- Rips, L. J. (1988). Deduction. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 116–152). Cambridge, UK: Cambridge University Press.
- Rogers, R. D., & Monsell, S. (1995). Costs of a predictable switch between simple cognitive tasks. *Journal of Experimental Psychology: General*, 124, 207–231. doi:10.1037/0096-3445.124.2.207
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological Review*, 84, 127–190. doi:10.1037/0033-295X.84.2.127
- Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, 119, 3–22. doi:10.1037/0033-2909.119.1.3
- Stanovich, K. E. (1999). *Who is rational? Studies of individual differences in reasoning*. Mahwah, NJ: Erlbaum.
- Stanovich, K. E. (2009). Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory. In J. St. B. T. Evans, & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55–88). Oxford, UK: Oxford University Press.
- Stanovich, K. E., & West, R. F. (1998). Cognitive ability and variation in selection task performance. *Thinking & Reasoning*, 4, 193–230.
- Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate. *Behavioral and Brain Sciences*, 23, 645–665. disc. 665–726.
- Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39, 1275–1289. doi:10.3758/s13421-011-0104-1
- Towse, J. N., & Neil, D. (1998). Analyzing human random generation behavior: A review of methods used and a computer program for describing performance. *Behavior Research Methods*, 30, 583–591. doi:10.3758/BF03209475
- Towse, J. N., & Valentine, J. D. (1997). Random generation of numbers: A search for underlying processes. *European Journal of Cognitive Psychology*, 9, 381–400.